

A study of generative large language model for medical research and healthcare

Year: 2023 | Citations: 361 | Authors: C.A.I. Peng, Xi Yang, Aokun Chen, Kaleb E. Smith, Nima M. Pournejatian

Abstract

There are enormous enthusiasm and concerns in applying large language models (LLMs) to healthcare. Yet current assumptions are based on general-purpose LLMs such as ChatGPT, which are not developed for medical use. This study develops a generative clinical LLM, GatorTronGPT, using 277 billion words of text including (1) 82 billion words of clinical text from 126 clinical departments and approximately 2 million patients at the University of Florida Health and (2) 195 billion words of diverse general English text. We train GatorTronGPT using a GPT-3 architecture with up to 20 billion parameters and evaluate its utility for biomedical natural language processing (NLP) and healthcare text generation. GatorTronGPT improves biomedical natural language processing. We apply GatorTronGPT to generate 20 billion words of synthetic text. Synthetic NLP models trained using synthetic text generated by GatorTronGPT outperform models trained using real-world clinical text. Physicians' Turing test using 1 (worst) to 9 (best) scale shows that there are no significant differences in linguistic readability ($p = 0.22$; 6.57 of GatorTronGPT compared with 6.93 of human) and clinical relevance ($p = 0.91$; 7.0 of GatorTronGPT compared with 6.97 of human) and that physicians cannot differentiate them ($p < 0.001$). This study provides insights into the opportunities and challenges of LLMs for medical research and healthcare.